DOCUMENT RESUME

ED 237 577 TM 830 847

AUTHOR Baker, Linda
TITLE Differences in the Standards Used by College Students to Evaluate Their Comprehension of Expository Prose.
SPONS AGENCY National Inst. of Education (ED), Washington, DC.
PUB DATE [83]
GRANT NIE-G-81-0100
NOTE 41p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Cognitive Processes; *Evaluation Criteria; Higher Education; Prose; Reading Comprehension; *Reading Strategies; *Self Evaluation (Individuals); Standards
IDENTIFIERS *Comprehension Monitoring; *Embedding Transformations

ABSTRACT

College students differing in verbal ability read and evaluated their comprehension of expository passages. Three different types of problems were embedded within the passages to provide opportunities for students to reveal the use of different standards of evaluation. Half of the subjects were informed that they should use three particular standards in order to identify the problems (lexical, external consistency, and internal consistency); the remaining subjects were not given specific information as to the standards they should use. All problems subjects identified were classified as to the type of standard they reflected. The classification scheme consisted of the three targeted standards plus syntax, propositional cohesiveness, structural cohesiveness, and informational completeness. Of particular concern were differences in the standards adopted by students receiving specific instructions and those receiving general instructions. Use of the lexical standard did not differ with instructional specificity, suggesting that students spontaneously evaluate their understanding of individual words. Students receiving general instructions rarely used the external and internal consistency standards, suggesting these are not criteria students typically adopt. Instead, they commented frequently on the structural cohesiveness of the passages. Students with higher verbal ability exhibited more frequent and more varied standard use than those with lower verbal ability. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED237577

Differences in the Standards Used by College Students to Evaluate their Comprehension of Expository Prose

Linda Baker

University of Maryland Baltimore County

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Running head: Evaluating Expository Prose

TM 830 847

Abstract

College students differing in verbal ability read and evaluated their comprehension of expository passages. Three different types of problems were embedded within the passages to provide opportunities for students to reveal the use of different standards of evaluation. Half of the subjects were informed that they should use three particular standards in order to identify the problems (lexical, external consistency, and internal consistency); the remaining subjects were not given specific information as to the standards they should use. All problems subjects identified, regardless of whether or not they were intentionally introduced, were classified as to the type of standard they reflected. The classification scheme consisted of the three targeted standards plus four others: syntax, propositional cohesiveness, structural cohesiveness, and informational completeness. Of particular concern were differences in the standards adopted by students receiving specific instructions and those receiving general instructions. Use of the lexical standard did not differ with instructional specificity, suggesting that students spontaneously evaluate their understanding of individual words. However, students receiving general instructions rarely used the external and internal consistency standards, suggesting these are not criteria students typically adopt. Instead, they commented frequently on the structural cohesiveness of the passages. Additionally, students with higher verbal ability exhibited more frequent and more varied standard use than those with lower verbal ability.

Imagine you are a college student reading a section in one of your textbooks and you are serious in your goal to understand the material. How would you decide whether or not you understand? That is, what criteria or standards would you use for evaluating your comprehension? There are probably some criteria that virtually all students would use, while there may be others that are adopted by only a few. For example, most students would probably decide they are having trouble understanding the material if they encounter many words whose meaning they do not know. Since the effect of vocabulary knowledge on comprehension is widely recognized, a lexical standard should be one which is frequently adopted. Suppose the passage contained information related to something you had previously learned. Would you consider how the new information relates to your prior knowledge? Or would you accept the information at face value, never questioning the truthfulness of the text? Unfortunately, the latter course may be more common, reflecting a failure to evaluate text for external consistency.

At the present time, there is little empirical evidence upon which to base an answer to the question posed above. That readers in fact evaluate their understanding is implicit in models of comprehension (e.g., Collins, Brown, & Larkin, 1980; Just & Carpenter, 1980) and explicit in models of comprehension monitoring (e.g., Baker & Brown, in press; Flavell, 1981; Lefebvre-Pinard & Pinard, in press; Markman, 1981). Nevertheless, theorists have been very vague as to how readers decide whether or not they understand. However, Baker (in press) has recently proposed that evaluation involves the application of standards which fall into three basic classes: lexical, syntactic, and semantic. The semantic class is further subdivided into

standards of external consistency, internal consistency, propositional cohesiveness, structural cohesiveness, and informational completeness. Evidence is cited from a wide variety of sources showing that readers are sensitive to manipulations that affect the comprehensibility of text along each of these dimensions.

The advantage of considering evaluation with respect to distinct dimensions becomes apparent when we examine the conclusions that have been drawn in research on comprehension monitoring. The typical paradigm has been to introduce a particular type of problem into a passage and assess a reader's ability to detect it. Consider, for example, a passage containing the following sentences: "The temperature on Venus is much higher than boiling water. But it is much too cold for us to live there." In order to realize that the ideas in these sentences are contradictory, the reader must evaluate the internal consistency of the text. Now consider the following: "States with many rotusts rely heavily on the sales tax." In order to realize that the meaning of a particular word is not known, the reader must apply a lexical standard of evaluation. Failures to notice both types of problems have been attributed to ineffective comprehension monitoring (e.g., Markman, 1979; Paris & Myers, 1981), even though detection of the problems requires that two different standards of evaluation be applied. In other words, the results reflect failures to use particular standards of evaluation; they reveal nothing about the possible use of alternative standards. It may well be that the same person who fails to use an internal consistency standard would be very likely to apply a lexical standard.

Several researchers have reported developmental and reader proficiency differences in comprehension monitoring using this global definition (e.g., Garner, 1980, 1981; Markman, 1979; Harris, Kruithof, Terwogt, & Visser, 1981; Paris & Myers, 1981; Winograd & Johnston, 1981). However, comprehension monitoring should not be regarded as a unitary phenomenon, something at which a reader is either "good" or "poor." If we are to help students evaluate their understanding more effectively, we need to acquire a better understanding of differences in the use of specific standards. The present study was designed to provide a foundation for this effort. It addressed such questions as whether some standards are more likely to be applied than others and whether more effective readers differ from less effective readers in either the scope of their standard repertoire (i.e., the number of different standards used) or the composition of the repertoire (i.e., the specific standards used).

College students differing in verbal ability participated in a study whose stated purpose was to find out what kinds of things make passages hard to understand. The passages, excerpted from college-level textbooks, contained many potential sources of difficulty, covering the full range of standards described by Baker (in press). In addition, three types of problems were intentionally embedded within the passages, problems which could be detected by using the lexical, external consistency, and internal consistency standards of evaluation. Half of the subjects within each verbal ability group were specifically told that these three types of problems would be present, and they were given examples of each type. The remaining subjects were simply told that problems would be present. Of particular interest here

was whether the subjects receiving general instructions would be less likely to apply any of the targeted standards than those receiving specific instructions and whether they would be more likely to spontaneously adopt any of the non-targeted standards. Also of concern were possible interactions with verbal ability level.

The primary focus of the paper will be on the use of different standards rather than on the identification of embedded problems. This focus reflects a shift from previous research which operationalized comprehension monitoring as success at detecting specific problems. There are several reasons for the shift. The first, mentioned earlier, is that failure to notice a specific type of problem reflects a failure to use a specific standard of evaluation, not a failure to monitor comprehension. A second reason is that failures to notice particular problems may be due to a number of different factors, including lack of relevant prior knowledge, misinterpretations, and inferential reasoning that serves to resolve the problems (cf. Baker, 1979). It seemed more sensible, therefore, to focus on anything a particular subject perceived as problematic, regardless of whether or not the experimenter perceived it as such. This approach affords a much richer perspective on the ways that readers evaluate their understanding. A third reason for the shift is that differences in detection of the different problem types are difficult to interpret (cf. Markman & Gorin, 1981). Since the problem types are qualitatively different from one another, we do not know whether they would be equally easy to detect if the appropriate standards were used effectively.

Method

Subjects

A total of 59 college students participated in the study, 37 women and 22 men. The students were enrolled in an introductory psychology course and participated in order to receive extra credit. The majority of the subjects were first-semester freshmen. In order to obtain two groups of subjects who differed considerably in verbal ability level, only those students whose SAT-Verbal scores fell below 420 and above 580 were asked to participate. The final sample consisted of 23 students with higher verbal ability (mean=599, s.d.=66) and 35 with lower (mean=388, s.d.=62)[1]. The inequality of group sizes was not apparent until some time after the data were collected; several students reported that their scores were in the high category, but in fact their scores were low. All subjects were asked at the outset of the experiment to sign a form giving permission for their SAT scores to be released from the admissions office. When the official scores were received, several students had to be reclassified. Additional data collected on each subject were the number of semesters enrolled in college, date of birth, final introductory psychology course grade, and Math SAT score.

[1] The sign-up sheet did not specify that only native speakers of English should sign up. Consequently, one non-native speaker did participate. His Verbal SAT score was 230, but his performance on the detection task was better than any other subject. Moreover, his Math SAT score was 650 and his final introductory psychology grade was an A. Since his Verbal score was clearly not representative of his abilities, this subject was dropped from the analyses.

Subjects were assigned to either the general or specific instruction condition when they reported to the laboratory. An attempt was made to keep the conditions as balanced as possible for verbal ability level and gender throughout the course of testing. In order to prevent possible bias in testing, the subjects were asked to wait until the end of the experiment to report their SAT scores. At the beginning of the session they were simply asked whether their score did meet the requirements specified on the subject sign-up sheet.

Materials

The materials consisted of 12 expository passages adapted from college-level textbooks. Four selections were taken from each of three different texts, one dealing with animal behavior, a second with state and local government, and a third with environmental/ecological issues. Examples of passage topics from each book, respectively, are: temperature regulation in lizards, advantages of the sales tax, and the disposal of solid wastes. The topics were selected to contain facts not generally known and not likely to be encountered in the beginning college student's coursework. The selections were shortened, if necessary, so that all of the passages would be roughly comparable in length. Each passage had 11 or 12 sentences; the mean number of words was 202 (s.d.=16.48). The readability of the passages was measured using Fry's (1978) scale; however, no attempt was made to modify any passage's readability level (mean grade level=11.67, s.d.=2.50).

The passages varied considerably in style and structure, as well as in the quality of writing, thus affording ample opportunities for the students to question various aspects of the passages. However, it was felt inappropriate

to rely exclusively on naturally-occurring opportunities to identify problematic sections of the passages; therefore nine of the passages were intentionally disrupted. The disruptions consisted of nonsense words, prior knowledge violations, and inconsistencies. These different problem types could be identified by effective use of the lexical, external consistency and internal consistency standards of evaluation, respectively. One of each problem type was present in each of the three content areas. All of the problems were created by replacing a single word with another word. The substitutions occurred in the middle of a sentence which appeared in any serial position from five through eight. Three of the modified passages contained two-syllable nonsense words which followed standard rules of English orthography. An example of a nonsense word appears in the following sentence: "The landfills must not be in colactins (locations) that could lead through seepage to contamination of ground water." Three other passages presented facts that violated common world knowledge. An example is: 'During the summer (winter) months most frogs hibernate in the soft mud of the bottom of pools and streams." The third type of problem, also embedded in three passages, was internal inconsistency, created by making one sentence in the passage conflict with a previously presented sentence. One sentence always intervened between these two target sentences. An example follows: "These small [housing] firms usually assemble the principal components of the house themselves. The reasons for the differences between these two industries [housing and auto] lie in the fact that local government is the primary regulator of land use in this country. Zoning and building code regulations are responsible for the presence (absence) of large mass-production oriented

construction firms." The remaining three passages did not contain any intentionally-introduced problems. A fifth selection from one of the texts served as a warm-up passage.

A preliminary screening study was carried out in which 23 advanced undergraduates were instructed to read the passages carefully and to comment on any aspects that seemed problematic. They were specifically told that problems were present and were given examples of the three types that had been introduced, but they were given no hint as to how many problems might be present in a given passage. The purpose of the study was threefold: 1) to ensure that there were neither ceiling nor floor effects in the detectability of any problems; 2) to ensure that there was not extensive variability in the detectability of the different exemplars of a given problem type; and 3) to verify that there would be other sections of the passages that individual subjects would identify as problematic. Subjects' responses were scored both for problem identification and for the other kinds of comments they made. Detection rates ranged from 20% to 80%, neither at ceiling nor floor. Inconsistencies and prior knowledge violations averaged 30%, nonsense words 70%. Within problem types, there was some but not extensive variability. Subjects did indeed identify many additional parts of the passages as problematic. For the most part, these additional comments were idiosyncratic, though occasionally two or three people had the same comment. Moreover, the comments reflected a full range of standard use; subjects did not restrict themselves to the three standards that were needed to detect the intended problems. Also present were comments about syntax, structural and propositional cohesiveness, and informational clarity and completeness.

The final versions of the passages were assembled into booklets. Each passage was typed on a separate sheet of paper and was headed by a descriptive title. All passages appeared as a single long paragraph, even if more paragraphs had been present in the original text. (Note that this gives subjects occasion to comment on the structural cohesiveness of the text.) At the bottom of each page was a set of four evaluative sentences which were to be used by the subjects to indicate whether the passage was problematic, and if so, to what extent did it affect comprehensibility (see Procedure section). The same passage appeared in first position in each booklet and served as a practice passage. The remaining 12 passages were arranged in a constrained, partially randomized, order. The constraint was that a particular problem type had to appear in a particular position across booklets, but the passage containing that problem was randomly selected from the pool. For example, the first experimental passage contained a prior knowledge violation; for any given subject it could be any one of the three prior knowledge passages. The second passage always contained an internal inconsistency, and it too could be any one of three, etc.

Procedure

All subjects were seen individually by the author. At the beginning of each session, the subjects were asked if they would be willing to authorize release of their official test scores and if so, to sign a consent form. (All subjects were willing.) They were then given a booklet containing the passages and were asked to fill in on the cover sheet their names, ages, dates of birth, and number of semesters they had attended college. At the bottom of the page were four statements that the subjects were asked to refer to while they listened to the relevant portions of the instructions.

The instructions were presented on tape, recorded by the experimenter. Subjects in both conditions were told that the purpose of the study was to try to find out what kinds of things make passages hard to understand. They were instructed that their task was to identify problems that had been deliberately introduced into some expository passages similar to those they might find in college textbooks. A problem was defined as "something that might confuse people or something that people might have trouble understanding." Subjects in the specific instruction condition were given further information about the exact nature of the problems and two examples of each type. The examples of the internal inconsistencies and prior knowledge violations were the same as those used by Markman and Gorin (1981); the nonsense word examples followed the same format. The terms used to describe the three problem types were: 1) two parts of the paragraph that don't make sense together; 2) things that aren't true; and 3) words that aren't really words. Subjects were not told how many problems might be present in any one passage.

Subjects in both conditions were instructed to underline anything they thought was a problem as they were reading. Then, when they finished, they were to check one of the statements at the bottom of the page in the space provided. The statements were: 1) There were no problems; 2) There may have been a problem but it did not make the passage hard to understand; 3) There was a problem and it made the passage a little hard to understand; and 4) There was a problem and it made the passage harder to understand. (As it turned out, there were large individual variations in the use of the rating scale; some subjects only used the first two statements, others used the full range. For this reason, the rating data were not subjected to any statistical

analyses.) The subjects were further informed that after they checked a statement they were to explain why they made that choice and also why they underlined any words or phrases.

After the instructions were presented, the subjects were asked to turn to the first (practice) passage and read it silently to themselves. Care was taken to ensure that the subjects understood the task by reminding them, if necessary, to underline problematic information as soon as they encountered it and by pointing out any lack of correspondence between ratings and underlining (e.g., statement #3 was checked but no sentences were underlined). The subjects were not given any feedback. Once it was clear that they understood the task, the subjects went on to the 12 experimental passages. For each passage, the subjects read it silently at their own pace, underlining any problematic sections as they went along. When they finished reading the passage, they checked one of the rating sentences, and then explained why they had done so and, if appropriate, why they underlined what they did.

Upon completion of the task, the subjects were asked a standard set of questions designed to reveal how they had interpreted any problems they did not report. Before asking each question, the experimenter turned to the appropriate passage and placed it in front of the subject, who was encouraged to reread it to find the answer. If the subject did not spontaneously indicate awareness of the problem after answering the question, the experimenter asked, "Does that make sense to you?" If the subject still did not indicate problem awareness, a second prepared question was asked. To illustrate, the questions for the prior knowledge problem embedded in the sentence, "During the summer months, most frogs hibernate in the soft mud..."

were as follows: When do frogs hibernate? and Do frogs typically hibernate in the summer?

All sessions were tape recorded, and the tapes were subsequently transcribed. The length of the sessions varied, averaging about 40 minutes per subject, with a range of 25 to 60 minutes.

Scoring

Problem identification. Responses were scored as problem identifications if the subject underlined the target information and gave an adequate explanation of the nature of the problem. It was not necessary for the subject to specify the problem type. So, for example, if a subject underlined the nonsense word colactins and then said, "I don't know what that word means," the response was scored as a correct identification. Similarly, if a subject explained that she underlined the phrase, "during the summer months," because she didn't think frogs hibernated in summer, this was scored as correct. There was never any question on the basis of the responses as to whether the subject noticed the intended problems.

Standard application. Once the response protocols were fully transcribed, the subjects' responses to "non-problematic" segments of text were coded as to the type of standard they revealed. The coding scheme used was that established by Baker (in press) and included seven categories: lexical, external consistency, internal consistency, syntax, informational clarity and completeness, propositional cohesiveness, and structural cohesiveness. Examples of comments reflecting the use of these standards are presented in Table 1, along with the relevant passage contexts.

Insert Table 1 about here

It is important to note that no judgments were made as to whether the information was in fact problematic. For example, if a subject misunderstood a particular sentence and so said that the following sentence was inconsistent, this was scored as an application of the internal consistency standard just as if the two sentences were in fact inconsistent. The protocols were scored by two independent judges who were able to classify all but 2% of the comments into one of the seven categories. Inter-judge agreement was nigh and the few discrepancies were resolved through discussion.

Results

This section of the paper is divided into several different sections. The first section presents the results for the problem identification task, reporting both an analysis of variance and a multiple regression analysis. The second, most important, section focuses on the application of the various standards as the subjects were reading the entire passages. Several different dependent measures will be examined: 1) frequency of use, 2) number of different standards used, and 3) presence or absence of particular standards in a subject's repertoire. The third section will briefly report the more salient aspects of subject's responses to the follow-up questions about problems they failed to report.

Problem Identification

The mean number of problems of each type that the students identified is shown in Table 2. Quite clearly, the students did not do very well on this

task. An analysis of variance with instruction condition and verbal ability level as between-subject factors and problem type as a within-subject factor revealed reliable main effects of each factor as well as an instruction by verbal ability interaction. Students who received specific instructions as to the types of problems they should seek identified more problems (44%) than those who were simply informed that problems would be present (25%), F (1, 54) = 20.85, $p < .001$. Additionally, students who scored higher on the Verbal port on of the SAT identified more problems (42%) than those who scored lower (27%), F (1, 54) = 13.02, $p < .001$. Interpretation of these main effects is qualified by the interaction, F (1, 54) = 5.06, $p < .05$: higher verbal ability students who received general instructions did not identify more problems than the lower verbal ability students receiving general instructions. Moreover, the performance of this higher ability group did not

---

Insert Table 2 about here

---

differ from that of the lower ability students who received specific instructions. Although specific instructions were associated with better problem identification for both higher and lower ability students, the magnitude of the difference between instruction types was much greater for the higher ability group.

The identification rates for the different problem types differed, as expected, F (2, 108) = 76.79, $p < .001$. Many more nonsense words were identified than either prior knowledge violations or inconsistencies (66% vs. 21% vs. 10%, respectively). The difference between the latter two problem

types was also statistically significant (Fisher's lsd = .28). There were no interactions of problem type with instructions or verbal ability (Fs < 1).

The above analysis of the data was based on a dichotomous classification of subjects into higher and lower verbal ability levels. The reliable main effect of verbal ability indicates that the two groups did indeed differ. But it does not indicate how much of the variance in problem identification is attributable to verbal ability. To answer this question, multiple regression analyses were carried out using the subject's actual Verbal SAT score as a predictor variable. Additional predictor variables were Math SAT score, Introductory Psychology final grade, age, and number of semesters enrolled in college. Because instruction condition was a qualitative variable, the data were analyzed separately for the two conditions. The total number of problems identified served as the dependent variable.

The regression analyses for the two different instruction conditions yielded very different outcomes, as might be expected given the analysis of variance results. We will consider first subjects in the specific instruction condition. The univariate correlations with problem identification for the Verbal SAT, Math SAT, and course grade, respectively, were .72, .67, and .39. However, when the effects of the Verbal score were partialled out, neither the Math score nor the course grade correlated reliably with problem identification. The predictor variables were entered into the regression equation through a forward stepping algorithm which required a minimum F-ratio of 4.0 for entry. Verbal score was the only variable to enter the regression equation (F to enter = 25.39), accounting for a substantial proportion of the variance (Multiple r^2 = .51). In contrast, under general instructions, verbal

ability had virtually no predictive value for problem identification. Univariate correlations for Verbal SAT, Math SAT and course grade, respectively, were .14, .36, and .03. None of the variables had sufficient F-levels to enter the regression equation, though the Math SAT score was close, with an F to enter of 3.91. Even this variable, however, accounted for but 10% of the variance.

Thus, the regression analyses lend further support to the conclusion drawn on the basis of the analysis of variance: whereas verbal ability level has a strong effect on students' ability to identify problems when they are specifically instructed what to look for, it has virtually no effect on problem identification when the problem types are not specified.

Application of Standards Throughout the Testing Session

As indicated in the Method section, the data to be considered in the following analyses are based on all problems that the students identified, regardless of whether they had been intentionally introduced by the experimenter. Moreover, these data are based on all 12 passages the subjects read, not just the 9 that contained intentional problems. Several different questions about standard use will be addressed. The first concerns the frequency with which subjects used each standard, the second concerns the number of different standards each subject used, and the third concerns patterns of standard use, i.e., whether there were some standards that subjects never used at all or used to the exclusion of any others.

Frequency of use. The mean number of times students applied each standard is presented in Table 3 as a function of instruction type and verbal ability level. Note that the standards differed considerably in their

frequency of use, as expected. Note also that the pattern differs depending on whether the students received general or specific instructions. A 2 (instruction) x 2 (verbal ability) x 7 (type of standard) mixed analysis of variance confirmed these apparent differences. The main effect of standard type was reliable, F (6, 324) = 21.66, $p < .001$, as was the standard by instruction interaction, F (6, 324) = 9.11, $p < .001$. In addition, the lower verbal ability group overall used the standards less frequently, F (1, 54) = 4.44, $p < .05$. There was also a tendency towards a three-way interaction, F (6, 324) = 1.81, $p < .10$.

Insert Table 3 about here

One might have predicted that subjects receiving specific instructions would apply the three instructed standards more frequently than subjects receiving general instructions, leading to a main effect of instruction condition. However, the main effect was not reliable, F (1, 54) = 1.27, $p > .10$, although there was the interaction with standard type. Fisher's lsd procedure was used to test the differences between means (lsd = .68, $p < .05$). Specific instructions did in fact lead to increased use of the external and internal consistency standards, but the lexical standard was used equally often by both groups of subjects. One of the noninstructed standards, structural cohesiveness, was used more often by subjects receiving general instructions. The remaining three standards were used with equal frequency by subjects in both groups.

Of particular interest are the types of standards the students adopted most frequently when simply given general instructions to find the problems, since this has implications about how students would spontaneously evaluate their understanding when motivated to process text carefully. The lexical standard and the structural cohesiveness standard were used more often than any others. Note that the mean for the lexical standard is greater than 3, which would be expected if students reported all of the nonsense words and nothing else. However, we know from the preceding section that students receiving general instructions identified fewer nonsense words than subjects receiving specific instructions; yet they used the lexical standard equally often. It seems, then, that they were particularly sensitive to vocabulary items whose meaning they were not sure of while students receiving specific instructions let pass words they recognized but perhaps could not define.

The high frequency of use of the structural cohesiveness standard indicates that not only is comprehension affected by text structure characteristics, as many have argued (e.g., Kieras, 1980; Kintsch & Yarbrough, 1982), but also that students are conscious of its effects. It is interesting to note that the lower verbal ability students used the standard considerably less often than the higher. This is in line with evidence that poorer readers tend to be less sensitive to text structure disruptions than better readers (e.g., Eamon, 1978/79; Meyer, Brandt, & Bluth, 1980).

Another standard that the general instruction subjects used relatively often was the informational clarity and completeness standard. It was used less often than the lexical or structural standards but more often than the others. Here again, the higher ability students tended to use the standard

more often than the lower ability students, suggesting greater concern with whether the information was sufficient to enable them to grasp the point the author was trying to make.

None of the other standards were used more than once during the entire session by the average subject. It is particularly noteworthy that the external consistency standard falls into this group. In a similar type of study involving fourth and sixth graders, Baker (Note 1) found that the external consistency standard was spontaneously adopted almost as often as the lexical standard. Other studies have also found that children are quite apt to challenge the truth of passage information (e.g., Markman, 1979). The present findings suggest that as students mature they become more likely to accept information as given, without critical evaluation of its truthfulness. However, that students will indeed apply the standard when instructed to do so is apparent by the large difference between subjects in the general and specific conditions.

We turn now to more detailed consideration of differences in standard use among the instructed subjects. The lexical and external consistency standards were used most frequently and equally often. However, the lower verbal ability students seem less likely to challenge passage truthfulness than the higher verbal students. Surprisingly, they also identified fewer word level problems. This may have indicated a reluctance to identify words as nonsense for fear that they might in fact be real words they should have known. The internal consistency standard was used less often than the other two, and here again, the lower verbal group used the standard less often than the higher verbal group. Somewhat unexpectedly, two other standards were used as

frequently as the internal consistency standard: structural cohesiveness and informational clarity/completeness. In other words, the subjects did not limit their comments to problems described in the instructions. The informational standard was used somewhat more often by the higher verbal students than the lower, while the opposite was true for the structural standard.

Number of different standards used. Of interest here was whether or not students in the different groups would differ in the variety of ways they evaluated their understanding. The number of different standards used at least once by each student was entered into a 2 (instruction) x 2 (verbal ability) analysis of variance. The analysis revealed a reliable main effect of verbal ability level, F (1, 54) = 3.98, p = .05. The higher verbal students used more different standards than did the lower verbal students (means = 4.21 and 3.53, respectively). This indicates that not only do the less effective readers report fewer obstacles to comprehension, as shown in the preceding sections, but they also have a more limited set of criteria against which to evaluate their understanding. The main effect of instruction type was not reliable (p = .14) nor was the interaction with verbal ability (F < 1). Thus, even those subjects who were not given any hint that there were three kinds of problems they should look for spontaneously applied an average of 3.61 different standards (as compared to 4.13 for the informed subjects).

Patterns of standard use. Another examination of group differences in the kinds of standards used was carried out by first classifying the subjects as to whether or not they ever used a particular standard and then entering the cell frequencies into multiway tables. Tests of association for patterns

of standard use were carried out using a log linear model. The relevant data are presented in Table 4 as proportions of subjects in each cell who used each standard (rather than as number of subjects who did or did not use each standard).

Insert Table 4 about here

We will consider the results for each standard in turn. With respect to the lexical standard, there were very few subjects in any condition who never used it; tests of association revealed no differences. The external consistency standard was adopted by most of the subjects in the specific instruction groups, but there were a majority in the general groups who did not use it, $\underline{X}^2$ (1) = 20.40, $\underline{p} < .001$. The internal consistency standard was also adopted by more subjects in the specific group than the general group, $\underline{X}^2$ (1) = 7.73, $\underline{p} < .01$. In addition, more higher verbal subjects used the standard than lower verbal, $\underline{X}^2$ (1) = 5.51, $\underline{p} < .05$. The two factors did not interact.

Fewer than half the subjects in any group adopted the propositional cohesiveness standard. None of the tests of association were reliable. The structural cohesiveness standard, in contrast, was adopted by most of the subjects in each group; again, the association tests were nonsignificant. Somewhat more than half the subjects adopted the informational completeness standard, with no differences in group frequency. Finally, with regard to the syntactic standard, virtually none of the subjects receiving specific instructions used this standard, while a quarter of those receiving general instructions did, $\underline{X}^2$ (1) = 7.03, $\underline{p} < .01$.

In summary, there were large differences among the standards in the likelihood that subjects would adopt them. The differences between instruction conditions in the use of the external and internal consistency standards are not surprising given the experimental manipulation. However, what is surprising is that the differences were so large. This suggests that a substantial proportion of adult readers never evaluate their understanding with respect to either of these two important criteria. Although subjects with general instructions were more likely to adopt the syntactic standard than subjects with specific instructions, the proportion was low relative to the other standards, indicating that these subjects too were typically more concerned with higher-level semantic features of the text. In particular, there were two semantic standards that most of these subjects adopted: structural cohesiveness and informational clarity/completeness.

The lexical standard, which was used by virtually all subjects, is not necessarily adaptive if it is used at the expense of higher-level standards. There is some evidence that younger and poorer readers in elementary school tend to rely exclusively on this standard (e.g., Baker, Note 1; Garner, 1981). It was of interest to ascertain whether less successful students at the college level also over-rely on this standard. We have already seen that students with lower verbal ability use fewer different standards than students with higher verbal ability. What standards did they in fact use? This question was answered by focusing on those subjects who used only one or two different standards. Single standard use was very rare, observed in only three subjects. Two of these subjects were in the lower ability group (one received general instructions, one specific) and the single standard used was

the lexical. (The third subject, in the higher ability-general group, used the structural cohesiveness standard exclusively.) There were seven subjects who used two standards only; for six of these subjects, the lexical standard was one of the two used. Four of these subjects were in the lower-general group, and two in the lower-specific. (The seventh subject, in the high-general group, used the structural cohesiveness and informational completeness standards.) Overall, 17% of the subjects had a repertoire limited to one or two standards. Though this is hardly a majority of the subjects, it is significant that 80% of this group had lower verbal ability and the same 80% always included the lexical standard in their repertoire. Thus, even among more mature readers, those who are less proficient show some overreliance on the lexical standard.

Responses to follow-up questions

The follow-up questions were included to reveal how students interpreted the intentionally-introduced problems they failed to report. By and large, the interpretations were those intended by the experimenter, indicating that the students failed to report the problems because they did not use the appropriate standard when it was needed. The typical comment, upon being confronted with a problem, was, "I can't believe I missed that!" There were a few exceptions, however, indicating not a failure to use a standard but rather a failure to perceive a problem. This was most apparent with one of the prior knowledge problems, embedded within the following sentence: "As governor of Montana, Ronald Reagan would typically invite legislators to his house..." Among the prior knowledge problems, this was in fact the one most frequently detected (p = .38). However, subsequent questioning of the non-detecting

subjects revealed that 24% of them were unaware that Reagan had not been governor of Montana. Moreover, all but one of these nine subjects were in the lower verbal ability group. Thus, not only are students of lower verbal ability less likely to evaluate text for consistency with prior knowledge, they also have a more impoverished knowledge base. Unfortunately, this latter problem is less amenable to instructional intervention than the former (Baker & Brown, in press). Failures to perceive the internal inconsistencies as problematic were also more characteristic of the lower verbal group than the higher. Here the difficulty is in logical reasoning, which again requires a different type of intervention than instruction designed to promote more effective evaluation of one's understanding. In short, the responses to the follow-up questions make it clear that the poorer problem identification by the lower ability students is not simply a product of less frequent or less effective use of the targeted standards of evaluation. More fundamental deficits in background knowledge and logical reasoning also play a role.

Discussion

In this section we will return to the question posed in the opening paragraph of the introcution about how a student would evaluate his or her understanding. The results of the present study offer many pieces of a comprehensive answer. As expected, it was found that readers do in fact use several different standards during a single interaction with a text. Even when given no hints as to what standards should be used, the average student employed 3.8 different standards. This indicates quite clearly that studies which examine the use of only a single standard may seriously underestimate the flexibility of students' evaluation activities. Nevertheless, students of

lower verbal ability did have more limited repertoires than their higher verbal ability peers. These limited repertoires tended to include the lexical standard, a more primitive standard that can be applied without regard to surrounding context. Although this standard should be included in any reader's repertoire (and was in fact used by 95% of the subjects), it should be used in concert with the higher-level semantic standards which take into account larger segments of the text.

The study also revealed that there were substantial differences in the frequency with which the different standards were used. Some differences would of course be expected among subjects who were told to use three specific standards. Of more interest were the standards used when students were not given any guidelines as to the kinds of things they should take into consideration. The lexical standard was very likely to be used by students whether or not they were instructed to use it, indicating a basic concern with the understanding of individual word meanings. However, students typically did not evaluate what they were reading with respect to what they already knew, even though they demonstrated use of this criterion when instructed to use it. Approximately two thirds of the subjects receiving general instructions never used the external consistency standard. Students tend to accept the facts presented in texts at face value, a tendency that automatically precludes any sort of critical reading. Interestingly, this tendency may increase with development, as suggested by evidence that children frequently challenge the truth of statements presented to them in expository passages even when they are not specifically instructed to do so (e.g., Baker, Note 1; Markman, 1979). Whether or not this apparent decline in the use of

the external consistency standard is a real phenomenon remains an important empirical question. If so, it may be an unfortunate by-product of instructional approaches stressing that the answer is in the book (Goodman, 1976). Another dimension unlikely to be considered spontaneously is the internal consistency of the text. Moreover, even when specifically instructed to evaluate for internal consistency, students with lower verbal skills did not use the standard very often. In fact, almost 50% of the subjects in this group never applied the standard. A number of component skills are involved in the evaluation of internal consistency, including the ability to retrieve and integrate nonadjacent propositions and draw logical inferences. These skills may well be limiting the lower verbal student's use of the standard.

We have seen, then, that students spontaneously adopted a lexical standard but seldom an external or internal consistency standard. However, they did use another standard as frequently as the lexical: structural cohesiveness. This indicates a concern with the thematic relationships among the ideas in the text and awareness of rhetorical conventions such as presenting topic sentences first. The standard is an important one for students to apply, especially when studying expository text, because it requires the identification of the main points of the passage. That 85% of the students receiving general instructions adopted this standard is encouraging; however, it is possible that they used it only because they could not think of any other kinds of problems that might be present and they knew they had to identify something. A third standard likely to be adopted spontaneously was informational completeness. Use of this standard reflects a concern with whether information is presented in sufficient depth or with

sufficient clarity to allow one to understand a particular concept. This standard is also important to use while studying because it indicates to the student whether there is a need to consult outside sources for more information.

Rarely used were the propositional cohesiveness or syntactic standards. However, the passages did not contain any intentionally-introduced problems which would elicit comments reflecting the use of these standards and there were not many naturally-occurring opportunities for their use. Therefore, the study does not provide clear answers as to the extent that these standards are typically used. We know that violations of propositional cohesiveness and syntactic constraints affect comprehension (e.g., Frazier & Rayner, 1982; Hirst & Brill, 1980), but it may be that such discrepancies are processed automatically and do not require the deliberate evocation of the relevant standards (cf. Baker, in press).

In conclusion, the study has revealed both quantitative and qualitative differences in the standards used by college students to evaluate their understanding of expository prose. It provides a rich source of hypotheses for further research designed to encourage students to adopt sensible and efficient criteria for evaluating their comprehension. For example, it illustrates the value of examining more systematically readers' skill at using all of the evaluation standards under conditions designed to elicit their use.

Reference Notes

1. Baker, L. Spontaneous versus instructed use of multiple standards for evaluating comprehension: Effects of age, reading proficiency, and type of standard. Unpublished manuscript, University of Maryland Baltimore County, 1983.

References

Baker, L. Comprehension monitoring: Identifying and coping with text confusions. Journal of Reading Behavior, 1979, 11, 363-374.

Baker, L. How do we know when we don't understand? Standards for evaluating text comprehension. In D. L. Forest, G. E. MacKinnon, & T. G. Waller (Eds.), Metacognition, cognition, and human performance. New York: Academic Press, in press.

Baker, L., & Brown, A. L. Metacognitive skills and reading. In P. D. Pearson (Ed.), Handbook of reading research. New York: Longman, in press.

Collins, A., Brown, J. S., & Larkin, K. M. Inference in text understanding. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, N.J.: Erlbaum, 1980.

Eamon, D. B. Selection and recall of topical information in prose by better and poorer readers. Reading Research Quarterly, 1978/79, 14, 243-257.

Flavell, J. H. Cognitive monitoring. In W. P. Dickson (Ed.), Children's oral communication skills. New York: Academic Press, 1981.

Frazier, L., & Rayner, K. Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. Cognitive Psychology, 1982, 14, 178-210.

Fry, E. B. Fry Readability Scale. Providence, R.I.: Jamestown, 1978.

Garner, R. Monitoring of understanding: An investigation of good and poor readers' awareness of induced miscomprehension of text. Journal of Reading Behavior, 1980, 12, 55-64.

Garner, R. Monitoring of passage inconsistency among poor comprehenders: A preliminary test of the "Piecemeal Processing" explanation. Journal of Educational Research, 1981, 74, 159-162.

Goodman, K. S. Behind the eye: What happens in reading. In H. Singer & R. B. Ruddell (Eds.), Theoretical models and processes of reading. Newark, Del.: International Reading Association, 1976.

Harris, P. L., Kruithof, A., Terwogt, M., & Visser, T. Children's detection and awareness of textual anomaly. Journal of Experimental Child Psychology, 1981, 31, 212-230.

Hirst, W., & Brill, G. A. Contextual aspects of pronoun assignment. Journal of Verbal Learning and Verbal Behavior, 1980, 19, 168-175.

Just, M. A., & Carpenter, P. A. A theory of reading: From eye fixations to comprehension. Psychological Review, 1980, 87, 329-354.

Kieras, D. E. Initial mention as a signal to thematic content in technical passages. Memory and Cognition, 1980, 8, 345-353.

Kintsch, W., & Yarbrough, J. C. Role of rhetorical structure in text comprehension. Journal of Educational Psychology, 1982, 74, 828-834.

Lefebvre-Pinard, M., & Pinard, A. Taking charge of one's cognitive activity: A moderator of competence. In E. Neimark (Ed.), Moderators of competence. Hillsdale, N.J.: Erlbaum, in press.

Markman, E. M. Realizing that you don't understand: Elementary school children's awareness of inconsistencies. Child Development, 1979, 50, 643-655.

Markman, E. M. Comprehension monitoring. In W. P. Dickson (Ed.), Children's oral communication skills. New York: Academic Press, 1981.

Markman, E. M., & Gorin, L. Children's ability to adjust their standards for evaluating comprehension. Journal of Educational Psychology, 1981, 73, 320-325.

Meyer, B. J. F., Brandt, D. M., & Bluth, G. J. Use of top-level structure in text: Key for reading comprehension of ninth-grade students. Reading Research Quarterly, 1980, 16, 72-103.

Paris, S. G., & Myers, M. Comprehension monitoring, memory, and study strategies of good and poor readers. Journal of Reading Behavior, 1981, 13, 5-22.

Winograd, P., & Johnston, P. Comprehension monitoring and the error detection paradigm. Journal of Reading Behavior, 1981, 14.

Footnotes

The research reported herein was supported in part by the National Institute of Education under grant NIE-G-81-0100.

Table 1

Examples of Comments Reflecting the Use of Different Standards of Evaluation

1. Lexical
   - A. Passage context: Some frogs defend themselves by feigning death.
     Comment: I don't think feigning is a word. (specific instructions-lower verbal)[a]
   - B. Passage context: Bats, the only group of flying mammals, are nocturnal insectivores.
     Comment: I don't know what insectivores are. (general instructions-higher verbal)

2. External Consistency
   - A. Passage context: Moves to get federal aid for the unemployed in the depressions of 1893-94, 1914, and 1921 failed.
     Comment: I don't know my history that well, but I don't think there were depressions all those times. (specific instructions-lower verbal)
   - B. Passage context: Franklin D. Roosevelt, elected president in 1933, immediately began to implement new relief programs.
     Comment: Elections are held in even-numbered years; 1933 can't be right. (general instructions-lower verbal) [This student, as well as 4 others, identified a problem embedded in the original text and unnoticed by the experimenter.]

3. Internal Consistency
   - A. Passage context: Filled landfill sites are sometimes used for golf courses and parks. But building on such sites is hazardous because of natural gas seepage from the underlying decaying garbage.
     Comment: Landfills are used for parks but they're too hazardous for houses? That's stupid. (specific instructions, higher verbal)
   - B. Passage context: Governors possess many formal powers to discharge the duties of chief executive, chief legislator, and head of state...

Governors frequently feel compelled to go to great lengths to win legislators over to their side.

Comment: It says governors have many formal powers, have rule over everyone. So why would they feel compelled to win legislators over to their side? (specific instructions-lower verbal)

4. Propositional Cohesiveness

   A. Passage context: Some [frogs] can blow up their lungs so that they are difficult to swallow.

      Comment: They are difficult to swallow. What are? (specific instructions-higher verbal)

   B. Passage context: Sales taxes are assessed on all those who buy taxable goods, regardless of their residence.

      Comment: Their residence. It almost sounds as though the things you're buying live somewhere else. (specific instructions-lower verbal)

5. Structural Cohesiveness

   A. Passage context: (The general topic of the passage is about the protective coloration of fish. The last two sentences appear here.) Some tropical fish are vividly colored, however, even in areas of dull and somber backgrounds. Although conspicuous in these circumstances, they are protected either by alertness and agility or by their poisonous flesh.

      Comment: Alertness and agility and poisonous flesh have nothing to do with color; the sentence is tacked onto the end. (general instructions-higher verbal)

   B. Passage context: (The passage is primarily concerned with the housing industry, but it begins with a comparison to the auto industry.)

      Comment: The title is Housing and Local Government. Then it goes into something about auto production which is totally irrelevant. (general instructions-lower verbal)

6. Informational Clarity and Completeness

   A. Passage context: (The general topic of the passage is pesticides and it includes a discussion of both herbicides and insecticides.)
   Comment: I know insecticides and herbicides are both types of pesticides and I know how they differ, but it's not specified in the passage; the terms should be defined more clearly. (specific instructions-higher verbal)

   B. Passage context. (The passage describes the special sense organs of snakes, including heat-sensitive pits.) The pits are supplied with a dense packing of free nerve endings from the fifth cranial nerve.
   Comment: The passage doesn't explain what the fifth cranial nerve is. (general instructions-lower verbal)

7. Syntax

   A. Passage context: The remaining 83 percent [of solid waste] is buried in landfill sites, with only about 10 percent of these being the so-called sanitary landfills.
   Comment: "With only about" is very awkward. (general instructions-lower verbal)

   B. Passage context: Seen from above against its normal background of water and stream bottom, the fish becomes almost invisible.
   Comment: "Seen from above" is hard to understand, there are three prepositions together; I had to read it three times. (general instructions-lower verbal)

a. Parenthetical remarks indicate the type of instructions and the verbal ability level of the subject making the comment.

Table 2

Mean Number of Problems Identified

		Problem Type		
Instructions	Verbal Ability	Nonsense Words	Prior Knowledge Violations	Internal Inconsistencies
Specific	Higher	2.75	1.17	1.08
	Lower	1.88	.75	.19
General	Higher	1.91	.45	.09
	Lower	1.63	.26	.05

Note: Maximum = 3.

Table 3

Mean Number of Times Subjects Applied Each Standard

	Instructions			
	Specific		General	
Type of Standard	Higher Verbal	Lower Verbal	Higher Verbal	Lower Verbal
Lexical	4.17	3.19	3.46	3.58
External Consistency	4.75	2.25	.82	.63
Internal Consistency	2.33	1.25	.45	.26
Propositional Cohesiveness	.58	.38	.82	.53
Structural Cohesiveness	1.00	2.00	4.00	2.95
Informational Completeness	2.17	1.00	1.82	1.00
Syntax	.17	.00	.36	.74
Total	15.17	10.07	11.73	9.69

Table 4

Proportion of Subjects Who
Applied a Particular Standard at Least Once

	Instructions			
	Specific		General	
Type of Standard	Higher Verbal	Lower Verbal	Higher Verbal	Lower Verbal
Lexical	1.00	.88	.91	.89
External Consistency	.92	.82	.36	.26
Internal Consistency	.92	.53	.45	.26
Propositional Cohesiveness	.42	.24	.36	.16
Structural Cohesiveness	.58	.82	.82	.89
Informational Completeness	.58	.53	.73	.58
Syntax	.08	.00	.27	.26